

Automatikus frázisdetektáló módszereken alapuló patológiás beszédelemzés magyar nyelven

Tündik Máté Ákos¹, Kiss Gábor¹, Sztahó Dávid¹, Szaszák György¹

Budapesti Műszaki és Gazdaságtudományi Egyetem,
Távközlési és Médiainformatikai Tanszék
e-mail: {tundik,kiss.gabor,sztaho,szaszak}@tmit.bme.hu

Kivonat A betegségek beszéd alapján történő korai diagnosztizálása során gyakori az automatikus osztályozási módszerek alkalmazása. Ezek az eljárások alkalmazhatók arra is, hogy Parkinson-kóros, valamint depressziós pácienseket az egészséges kontrollcsoport tagjaitól megkülönböztessük. A patológiás beszéd elemzése több szinten is elvégezhető; ebben a cikkben olyan prozódiai jellemzőket vizsgáltunk meg, melyek kinyerése automatikus hangsúly- és frázisdetektáló rendszerekből történt meg. Hipotéziseinket a frázisok időtartama és a frázisok szószámossága kapcsán fogalmaztuk meg. Az egészséges kontrollcsoport és a páciensek csoportja nagy mértékben elkülöníthető egymástól ezen paraméterek segítségével, melyeket statisztikai próbákból és SVM-alapú bináris osztályozásból származó eredményekkel alátámasztva az alábbi cikkben mutatunk be.

Kulcsszavak: gépi beszédfelismerés, gépi beszédértelmezés, hangsúly, frázis, Parkinson-kór, depresszió

1. Bevezetés

A betegségek beszéd alapján történő korai diagnosztizálása fontos kutatási terület, melyhez leginkább automatikus osztályozási módszereket alkalmaznak. Megvalósításukra számos példát találunk a szakirodalomban, pl. Bayes-típusú, SVM, mélyneurális háló, Random Forest, k-NN valamint GMM alapú [1] [2] [13]. A Parkinson-kór, valamint a depresszió esetén is cél, hogy a pácienseket az egészséges kontrollcsoport tagjaitól megkülönböztessük.

A Parkinson-kór az egyik leggyakoribb neurodegeneratív betegség, melynek a prevalenciája körülbelül 20/100 000 [7], előfordulása az életkor növekedésével egyenesen arányos. A betegség fő oka az agy feketeállományában lévő dopamintermelő idegsejtek nagymértékű sérülése, elhalálása. A Parkinson-kór fő tünetei közé tartozik a remegés, az izommerevség és a kognitív károsodás. Egyéb tünetei közé tartozik pl. a depresszió, demencia, alvászavarok. Friss kutatások szerint lehetőség van a Parkinson-kór beszéd alapú detektálására [6] [8]. A legtöbb beteg esetén beszédzavarok is fellépnek (diszfónia, dizartria), valamint a beszéd minőségében is változás lép fel. Jellemzővé válik a csökkent hangerő, a fokozott hangremegés és a levegősség [3] [8].

Hasonlóképpen, lehetőség van a depresszió beszéd alapú vizsgálatára is. Ismertetőjelei közé tartozik a lassú beszédtempó és a monoton beszédhang. Ezek a jellemzők konkrét akusztikai paraméterekhez köthetők, melyek szakirodalmi példákkal alátámaszthatóak. A hangulati, érzelmi állapotingadozás prozódiai paraméterekkel kapcsolható össze, mint pl. a ritmus, hanglejtés, hangsúly és az időzítés [15]. További vizsgált jellemzők lehetnek pl. az alaphangfrekvencia, a formánsok, a spektrális teljesítménysűrűség, kepsztrális vagy MFC együtthatók is [11] [12].

A [13] cikk szerzői a jitter, shimmer, HNR, alaphang jellemzőket a diszfónia különböző osztályainak megkülönböztetésére használták, majd ezeket összefüggésbe hozták a Parkinson-kór súlyosságának megállapításához használt UPDRS-értékekkel. Az osztályozás történhet kitartott hangok, olvasott szöveg, és spontán beszédfelvételek alapján. Az említett [13] cikkben 97%-os osztályozási teljesítményről számoltak be; Parkinson-kóros és egészséges csoportba történő bináris osztályozást végeztek, kitartott magánhangzók alapján. Egy másik cikk 85%-os pontosságról számol be; a beszéd érthetőségét folyamatos szöveggel vizsgálták, és "egészséges-enyhe-súlyos" Parkinson-kór osztályokat különböztettek meg [4].

A BME-TMIT Beszédakusztikai Laboratóriumában mind a Parkinson-kór, mind a depresszió beszéd alapú automatikus detektálása napjainkban is aktív kutatási terület [5] [10]. Jelen cikkben olyan prozódiai vonatkozású beszédjellemzőkre koncentráltunk, melyek automatikus hangsúly- és frázisdetektáló rendszerekből származnak. Feltételezésünk szerint az egészséges kontrollcsoport és a páciensek csoportja jól elkülöníthető egymástól ezen jellemzők segítségével, melyhez kapcsolódó kutatási eredményeinket az alábbi cikkben mutatjuk be.

A cikkünk az alábbi struktúra szerint épül fel: elsőként bemutatjuk a korpuszt, illetve a felhasznált automatikus hangsúly- és frázisdetektáló rendszereket. Ezután sor kerül a megkülönböztetési vizsgálatokhoz használt szempontrendszer leírására, végül ismertetjük az eredményeket.

2. Anyag és módszer

2.1. A felhasznált adatbázis

A kutatáshoz magyar nyelvű olvasott szöveget használtunk. Az egészséges kontrollcsoport résztvevői, valamint a depressziós és a Parkinson-kóros páciensek az "Az Északi szél és a Nap" c. történetet olvasták fel. Az említett történet egy fonetikailag kiegyensúlyozott rövid népmese (kb. hat mondat hosszú, és átlagosan 45 másodperc), gyakorlati jelentősége a foniátriai alkalmazásoknál van. Mivel a Parkinson-kóros és a depressziós páciensek különböző szövegezésű mesefordítást olvastak fel, ezért külön-külön egészséges kontrollcsoport összeállítására volt szükség annak érdekében, hogy a szövegezésből eredő különbségek méréseinket ne befolyásolják (a szöveg, illetve annak tagolása ugyanis nyilvánvalóan befolyásolja a prozódia). A 36 fős Parkinson-os csoporthoz 32, az 52 fős depressziós csoporthoz 36 fős egészséges csoport tartozik. Mind a kontrollcsoport, mind a páciensek felvételei csendes környezetben készültek el.

A kísérleteinkhez szó-, szótag- valamint fonéma szintű szegmentálások álltak rendelkezésre, ezeket kényszerített illesztéssel készítettük el, utólagos kézi korrekcióval. Ezen szegmentálások alapján lehetséges a fonológiai frázisok automatikus detektálása, melyet a 2.2 és 2.3 fejezetekben ismertetünk. A fonológiai frázisok prozódiai egységet képeznek, saját hangsúllyal és intonációs görbével jellemezhetők. A prozódiai hierarchiában elfoglalt pozíciójuk szerint több fonológiai frázis egy intonációs frázist alkot. A magyar nyelvre a kötött hangsúlyozás jellemző - a szavak első szótagján -, és a fonológiai frázis - definíció szerint - pontosan egy hangsúlyos elemet tartalmaz. A kézi fonológiai frázisszegmentálások elkészítésétől eltekintettünk, csak az automatikus módszerek segítségével kinyert frázisok alapján végeztük el a kiértékelést.

2.2. GMM/HMM alapú automatikus frázisdetektáló rendszer

Ebben a fejezetben egy GMM/HMM alapú automatikus frázisdetektáló rendszert mutatunk be. A módszer hét különböző fonológiai frázis modellezésére és detektálására alkalmas, gépi tanulás segítségével. A felügyelt tanulás miatt címkézett tanítóanyag szükséges. A tanításhoz 11 állapotú HMM/GMM modelleket használtak, az akusztikai-prozódiai jellemzők közül az alaphangfrekvenciára és a széles sávú energiára volt szükség. A modellek tanulásánál az akusztikai jellemzők első és másodrendű deriváltjait is felhasználták.

A megnyilatkozást tartalmazó beszédfelvétel fonológiai frázisokra történő szegmentálása a Viterbi-algoritmus segítségével történik meg. Mivel a magyar hangsúlyozási szabályok szerint a hangsúly az első szótagra esik, a fonológiai frázisok és a hangsúly detektálása közel egyszerre történik meg. A Viterbi-algoritmus mindegyik fonológiai frázis előfordulásához azonos valószínűséget rendel. A fonológiai frázisok, frázisintervallumok sűrűségét külön paraméterrel kontrollálhatjuk, melynek következménye pl. a megnövekvő hamis-pozitív elemek száma. Bővebb leírás a módszerről az alábbi [9] cikkben található.

2.3. WCAD intonációs modell alapú automatikus frázisdetektáló rendszer

Ebben a fejezetben egy intonációs modellezési technikán alapuló automatikus frázisdetektáló rendszert mutatunk be. A WCAD (Weighted Correlation based Atom Decomposition) rövidítés magyar megfelelője a súlyozott korreláción alapuló atom dekompozíciós algoritmus, mely fiziológiai alapokon nyugszik (hangszalagokat mozgató izmok feszítettsége), nyelvileg releváns információt hordozva. Az intonációs görbe rekonstrukciójának elméleti háttere a Fujisaki-modellből ered, mely a végleges kontúrt egy alap frekvenciakomponens, egy globális frázis "atom" valamint lokális "atomok" szuperpozíciójaként értelmezi.

A kísérleteinkhez kulcsfontosságú a lokális atomok időpontjának és amplitúdójának kinyerése. Az atomokat "Csúcs" és "Völgy" csoportra osztjuk, attól függően, hogy az adott szótagon jeleznek hangsúlyt, vagy a rákövetkezőn. A hangsúlyos szótagok kinyerése után a fonológiai frázisszegmentálás a magyar

nyelvre érvényes hangsúlyozási szabályok segítségével elkészíthető. A frázis kezdetén mindig hangsúlyos szótag ("Csúccsal" címkézve) áll, a végén pedig vagy "Völgygel" címkézett áll, ha a megnyilatkozásnak is vége, vagy pedig "Csúcs"-os, ha főmondat végén emelkedés következik be az intonációs görbében. A 2.2. fejezetben szereplő algoritmustól eltérően, az intervallumok bejelölése közvetlenül a beszédjelből kinyert jellemzők segítségével történik meg. Az egyes felvételekhez tartozó lokális atomok száma paraméterrel kontrollálható, mely kihat a fonológiai frázisokat tartalmazó intervallumok sűrűségére. A módszerről és annak hatékonyságáról részletesen a [14] számol be.

3. Az eredmények ismertetése

3.1. Hipotézisek

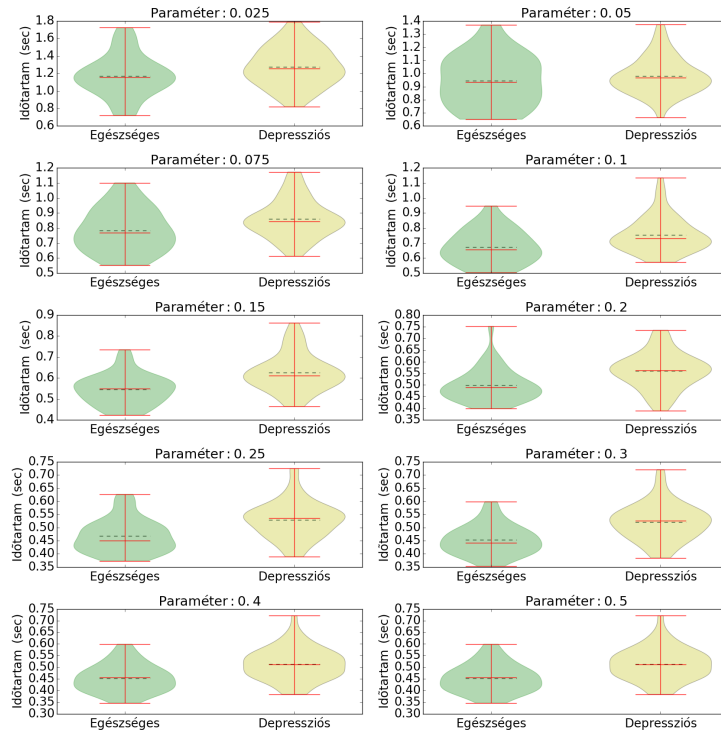
Korábbi vizsgálataink esetén egészséges emberek felvételeihez kézzel készített fonológiai frázis alapú szegmentációt használtunk fel, az automatikus módszerekkel történő összevetésre, az információelméletben használt fedés-pontosság (recall-precision) és F-mérték kiértékelési terében [14]. Ezen cikkben részletesen megvizsgáljuk az egészséges és beteg emberek közötti különbségeket a fonológiai frázisok aspektusában, de csakis az automatikus módszerek kimeneteire támaszkodva. Az alábbi szempontrendszert/hipotéziseket határozzuk meg:

- Frázisok időtartama - Azt feltételezzük, hogy az időtartam megkülönböztető jegy a betegek és az egészséges csoport között, mindkét frázisszegmentáló módszert alkalmazva.
- Frázisok szószámossága - Azt feltételezzük, hogy a frázisok hossza a benne lévő szavak számát tekintve megkülönböztető jegy a betegek és az egészséges csoport között, mindkét frázisszegmentáló módszert alkalmazva.
- t-próba és automatikus osztályozás az előző hipotézisekhez kapcsolódó eredményekre támaszkodva - Azt feltételezzük, hogy a frázisok időtartamának és szószámosságának megkülönböztető szerepe statisztikai úton is kimutatható, így ezek a jellemzők bináris osztályozásra is hatékonyan alkalmazhatók.

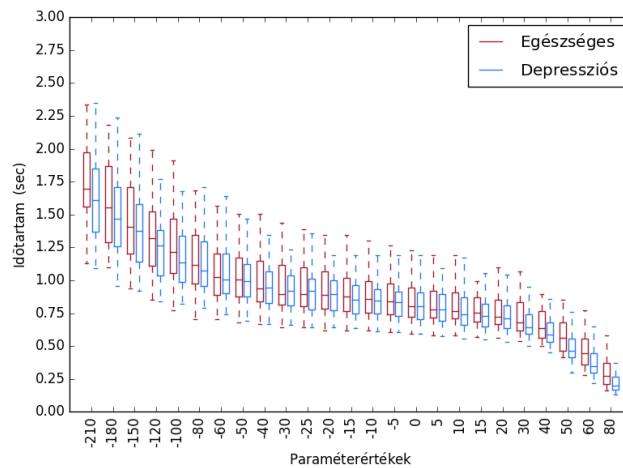
3.2. Frázisidőtartam elemzése

Az egyes frázisok időtartamának kinyerése a fonológiai frázisszegmentálás utófeldolgozó lépéseként könnyen elérhető. Ezeket mindkét módszer esetén kigyűjtöttük (a szünettel jelölt szakaszok kivételével), minden mérési szint esetén, melyek az algoritmusok paraméterkonfigurációból adódnak. Az eredményeket Violin Plot-okon illetve egyszerű Box Plot-okon foglaltuk össze.

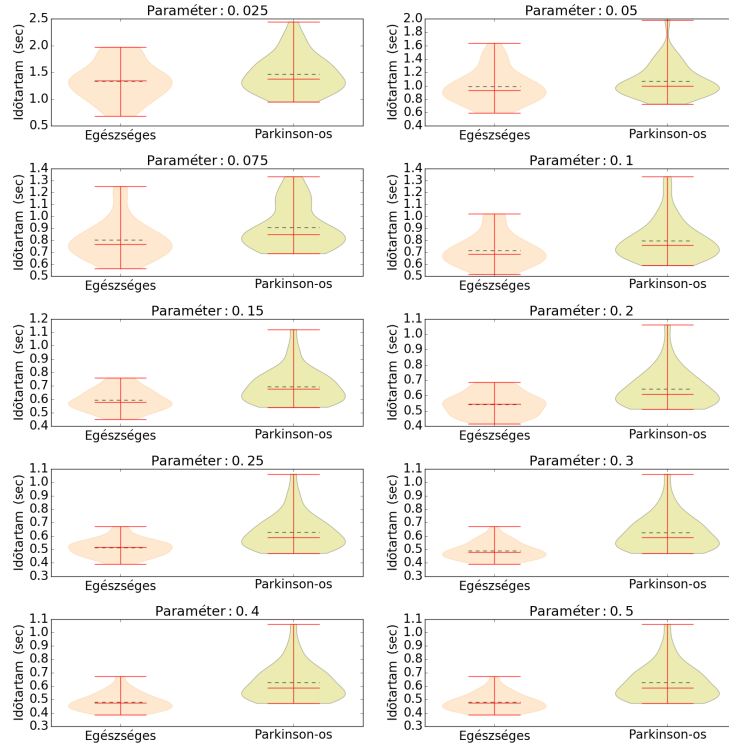
Az első megfigyelésünk az 1. és a 3. ábra alapján, hogy a frázisok időtartama egyre rövidebb lesz a WCAD-algoritmushoz tartozó paramétert magasabb értékre állítva (több frázishatároló kerül beszúrásra), valamint az egészséges csoport frázisai minden esetben rövidebbek, mint a Parkinson-os és depressziós csoporté.



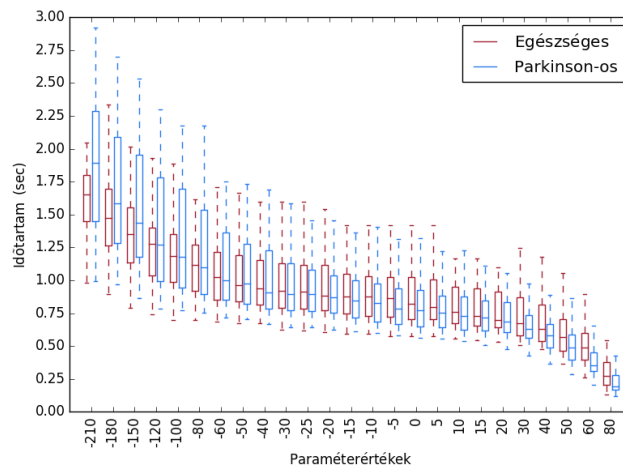
1. ábra. Frázisidőtartamok Violin Plot-ja, depressziós és egészséges csoportokra, WCAD-alapú automatikus frázisszegmentáló módszerrel



2. ábra. Frázisidőtartamok Box Plot-ja, depressziós és egészséges csoportokra, HMM-alapú automatikus frázisszegmentáló módszerrel



3. ábra. Frázisidőtartamok Violin Plot-ja, Parkinson-kóros és egészséges csoportokra, WCAD-alapú automatikus frázisszegmentáló módszerrel

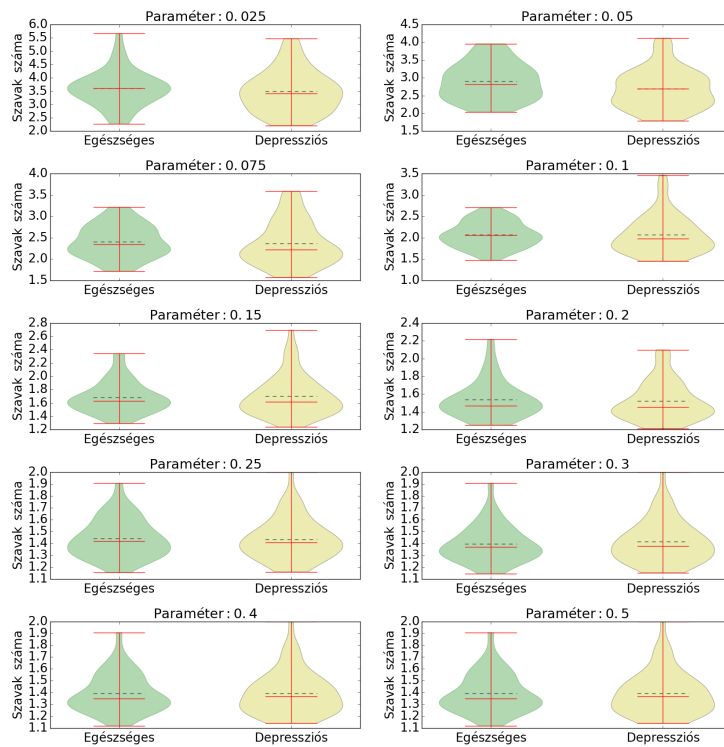


4. ábra. Frázisidőtartamok Box Plot-ja, Parkinson-kóros és egészséges csoportokra, HMM-alapú automatikus szegmentáló módszerrel

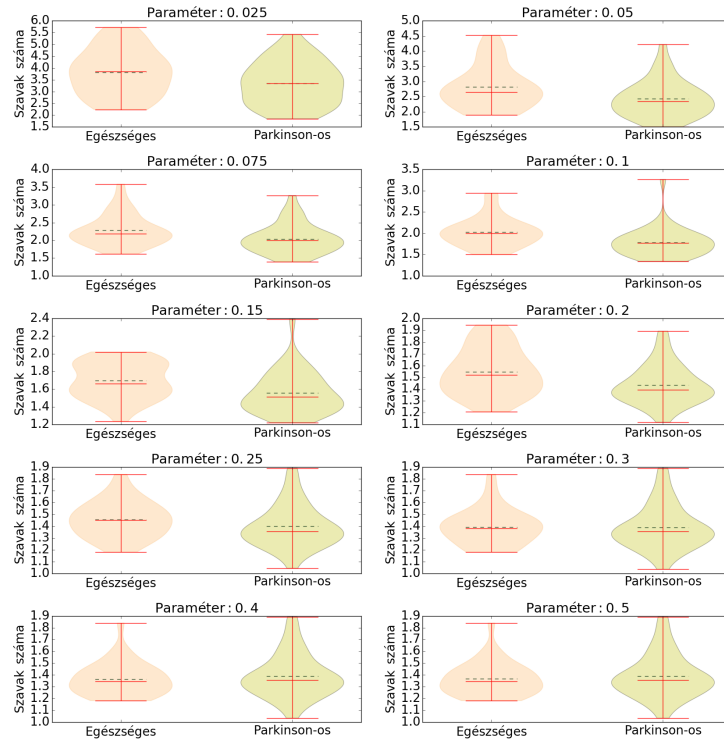
A következő megfigyelés a 2. és a 4. ábra alapján, hogy a frázisok időtartama a HMM-módszerhez tartozó "insertion log-likelihood" paraméter növelésével is csökken. Fontos különbség az előző módszer eredményéhez képest, hogy az egészséges csoportnál csak a mérési szintek első harmadában rövidebbek a frázisok, mint a Parkinson-osoknál, a depressziós csoporttal összehasonlítva pedig legtöbb esetben megegyeznek az értékek.

3.3. Frázisok szószámossága

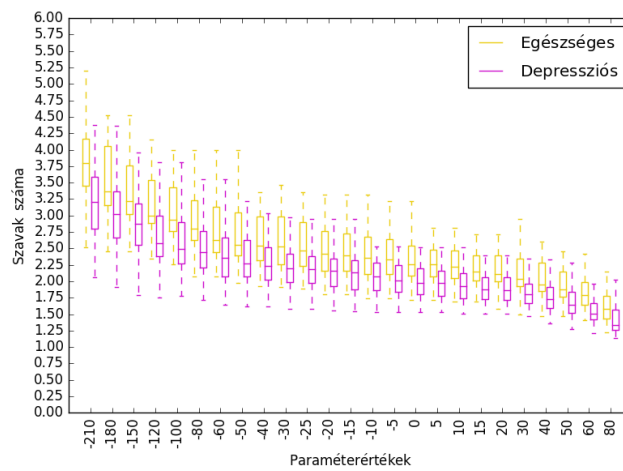
A frázisok szószintű kinyerése szintén utófeldolgozási feladat volt. A WCAD-alapú módszer során felhasználjuk a szótagok konkrét időpontját is. Így az egyes szavak időbeli illesztése az egyes frázishatárok közé egyszerűbb feladat, hiszen a frázis mindenképpen egy szó első szótagjával indul, valamint egy szó utolsó szótagjával végződik.



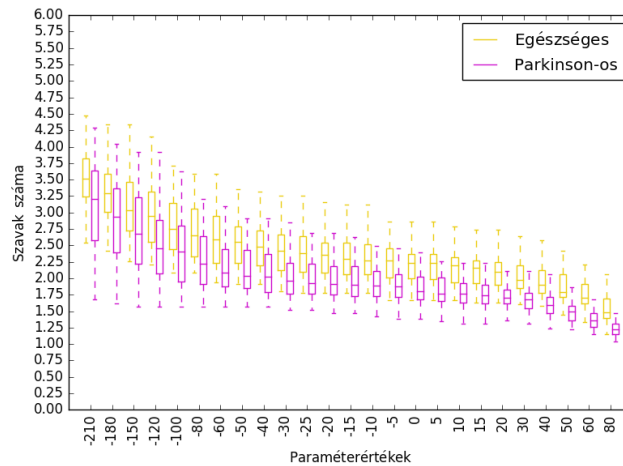
5. ábra. Frázisok átlagos szószámosságának Violin Plot-ja, depressziós és egészséges csoportokra, WCAD-alapú automatikus frázisszegmentáló módszerrel



6. ábra. Frázisok átlagos szószámosságának Violin Plot-ja, Parkinson-kóros és egészséges csoportokra, WCAD-alapú automatikus frázisszegmentáló módszerrel



7. ábra. Frázisok átlagos szószámosságának Box Plot-ja, depressziós és egészséges csoportokra, HMM-alapú automatikus frázisszegmentáló módszerrel



8. ábra. Frázisok átlagos szószámosságának Box Plot-ja, Parkinson-kóros és egészséges csoportokra, HMM-alapú automatikus frázisszegmentáló módszerrel

Más a helyzet a HMM-alapú rendszer esetén, ugyanis a rendszer nem használ szótag-információt, ezért a szavak frázisokra történő időbeli illesztésénél figyelnünk kell az időben átfedő szakaszokra is. Ezenkívül egyes frázisok közepén szünet is szerepel, mely pl. kézi szegmentációval nem fordulhatna elő; ilyenkor a szünetek mentén szét kell bontani a frázist, hogy a prozódiai hierarchiának megfelelő tagolást kapjunk.

A fent leírtak után elvégeztük a frázisok szószámosságának megállapítását, összesítve az egyes szakaszokban lévő szavakat, mindkét módszer mindegyik mérési szintjére, melyeket Violin Plot-okkal és Box Plot-okkal ábrázoltunk ismét.

Az első megfigyelésünk az 5. és a 6. ábrákkal kapcsolatban, hogy az egyes frázisokra kevesebb szó jut, a WCAD-alapú algoritmus paraméterét növelve (ez összefüggésben van az időbeli jellemzőknél leírtakkal). Az egészséges csoport frázisszintű megnyilatkozásai csak a mérési pontok első felében rövidebbek, mint a Parkinson-os csoporté, a depressziós csoporttal összevetve pedig közel azonosak.

Hasonlóképpen a 7. és a 8. ábrákon látható, hogy a HMM-alapú módszer paraméterét növelve a frázishosszok rövidebbek lesznek általánosságban, viszont a megfigyelések különbözőek; az egészséges csoportot összevetve a Parkinson-os és a Depressziós csoporttal minden mérési szint esetén rövidebb frázishosszokkal találkozunk.

3.4. t-próba és automatikus osztályozás

Az előző alfejezetben leíró jellegű statisztikai eredményekről számoltunk be, melyek a frázisok időtartamához és a bennük foglalt szavak számosságához kapcsolódtak. Következő lépésben a t-próba elvégzésével megvizsgáltuk, hogy a jellemzők esetén megfigyelt különbségek szignifikánsak-e. A frázisok időtartamát és

szószámosságát a frázisszegmentáló módszerek különböző paraméterértékeinek beállításával vizsgáltuk. WCAD alapú esetben 10, HMM alapú esetben pedig 25 különböző értéket használtunk. A H_0 hipotézis szerint az adott jellemzőnek nincs megkülönböztető szerepe az aktuálisan vizsgált módszer paraméterbeállítása mellett, ellenkező esetben a H_1 teljesül.

1. táblázat. Megkülönböztetés frázisidőtartam és a frázisokban lévő átlagos szószámosság alapján, mindkét frázisszegmentáló módszerrel

	t-próba eredmények($\alpha = 0.05$)		
	H_0	H_1	H1-paraméterek
Frázisidőtartam, Parkinson, WCAD	1	9	0.05—0.5
Frázisidőtartam, Depresszió, WCAD	1	9	0.025, 0.075—0.5
Frázisok szószámossága, Parkinson, WCAD	5	5	0.05—0.2
Frázisok szószámossága, Depresszió, WCAD	10	0	—
Frázisidőtartam, Parkinson, HMM	22	3	50—80
Frázisidőtartam, Depresszió, HMM	22	3	50—80
Frázisok szószámossága, Parkinson, HMM	0	25	-210—80
Frázisok szószámossága, Depresszió, HMM	0	25	-210—80

Az 1. táblázatból leolvasható, hogy a WCAD-alapú algoritmus az egészséges kontrollcsoportot és a pácienseket a frázisok időtartamát tekintve képes megkülönböztetni. Ezen kívül a Parkinson-os csoporttal összevetve a frázisokban szereplő szavak száma is megkülönböztető erővel bír, habár ez csak a mérési szintek felében igaz. Összehasonlításképpen, a HMM-alapú megoldás megkülönböztető ereje a frázisokban lévő szavak számában rejlik.

A következőkben az osztályzási kísérleteket és azok eredményeit ismertetjük. Minden kísérlet esetén az SVM C-SVC típusát használtuk, lineáris kernellel. A C hiperparaméter kimerítő kereséssel lett megállapítva, 2 első tíz hatványának szisztematikus végigpróbálgatásával. A tanítás és a tesztelés során „leave-one-out” keresztvalidációt alkalmaztunk.

2. táblázat. Eredmények SVM-alapú osztályzással

	SVM-eredmények	
	Pontosság (Acc)	C
Parkinson, WCAD	70,6%	8
Depresszió, WCAD	67 %	2
Parkinson, HMM	79,4%	2
Depresszió, HMM	62,5%	4
Parkinson, HMM+WCAD	88%	1
Depresszió, HMM+WCAD	81%	16
Parkinson, HMM+WCAD+FFS	91%	1
Depresszió, HMM+WCAD+FFS	83 %	16

Az eredmények a 2. táblázatban láthatók. Ha az egyes módszerekből származó jellemzőket külön tekintjük, a bináris osztályzás a Parkinson-Egészséges kontrollcsoport esetben a legpontosabb, a HMM-alapú fonológiai frázisszegmentálás jellemzőivel. Ha a két módszerből származó jellemzőket egyesítjük, további jelentős javulást érünk el, mind a "depressziós", mind a "Parkinson-os" esetben.

Végül, a Fast Forward Selection (FFS)-alapú jellemzőkiválasztással kaptuk a legjobb eredményeket, amelynek során minden egyes lépésben azt a jellemzőt választjuk ki az összességből, amellyel a legnagyobb mértékű javulás következik be a pontosság (accuracy) értékében. Depresszió esetén közel 40%-kal csökkent a dimenziók száma (37-ről 22-re), az eltávolított paraméterek főként a HMM-es algoritmusból származtak. Parkinson-os esetben még ennél is nagyobb, közel 90%-os dimenziócsökkenés következett be (42-ről 5-re), a megmaradt jellemzők többsége a WCAD-algoritmusból származik.

4. Összegzés

Cikkünkben részletesen bemutattuk az egészséges és Parkinson-kóros, valamint depressziós páciensek közötti különbségeket a fonológiai frázisok aspektusában, automatikus módszerek kimeneteire támaszkodva. Mivel a fonológiai frázisok egyszerre jellemzik a hangsúlyozást, az intonációt, illetve a prozódiai tagolást, lényegében ezen jellemzők tekintetében informatívak a kapott eredmények. Hipotéziseinket statisztikai vizsgálatokkal, valamint automatikus osztályozási kísérletekkel támasztottuk alá. Beigazolódott, hogy az automatikusan kinyert frázisok időtartama, valamint a frázisokra eső szószámosság megkülönböztető jegyek a betegek és az egészséges csoport között.

Az eredményeket értelmezve Parkinson-kóros felvételeknél a megakadások, szókeresések miatt kevesebb szó alkot egy frázist, a frázisok azonban időben így is hosszabbak, mint az egészségeseknél. Depressziósok esetén a frázisok nyúlnak, de kb. azonos szószámosságúak, mint az egészségeseknél. Néhány realizáció összehasonlítása alapján feltételezzük, hogy a monotonitás ellenére a frázishatárokat a rendszer többnyire ugyanott detektálja (nem lesz több szó a frázisban átlagosan), viszont a kisebb hangsúlyú, lapos intonációjú frázistípusok részaránya megnövekszik. Ezt a feltételezést a jövőben további kísérletekkel tervezzük igazolni. A legcélravezetőbb a HMM- és WCAD-jellemzők kombinált alkalmazása volt. A Parkinson-kór elkülönítése pontosabb a prozódia alapján (91%), ugyanakkor a depresszió esetén kapott 83% is kielégítőnek tekinthető.

További terveink között szerepel a magyar nyelvű adatbázis bővítése (a kísérletek elvégzése nagyobb elemszámmal), valamint egyéb jellemzőkkel történő együttes osztályzás és idegen nyelvű felvételek vizsgálata.

5. Köszönetnyilvánítás

A szerzők köszönetüket fejezik ki a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatalnak, amely a PD-112598 projekt keretében a kutatást támogatta.

Hivatkozások

1. Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O.: Collection and Analysis of a Parkinson Speech Dataset with Multiple Types of Sound Recordings. in *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, 828-834, (2013).
2. Frid, A., Safra, E.J., Hazan, H., Lokey, L.L., Hilu, D., Manevitz, L., Ramig, L.O.; Sapir, S.: Computational Diagnosis of Parkinson's Disease Directly from Natural Speech Using Machine Learning Techniques. in *Proc. of the 2014 IEEE International Conference on Software Science, Technology and Engineering*, 50-53, (2014).
3. Harel, B., Cannizzaro, M., Snyder, P. J.: Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. in *Brain and Cognition*, vol. 56, 24-29, (2004).
4. Khan, T., Westin, J., Dougherty, M.: Classification of speech intelligibility in Parkinson's disease. in *Biocybernetics and Biomedical Engineering*, vol. 34, Issue 1, 35-45, (2014).
5. Kiss, G., Tulics, M. G., Sztahó, D., Esposito, A., Vicsi, K.: Language Independent Detection Possibilities of Depression by speech. in *Recent Advances in Nonlinear Speech Processing*, 103-114, (2016).
6. Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., Ramig, L. O.: Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. in *IEEE Transactions On Biomedical Engineering*, vol. 56, no. 4, 1015-1022, (2009).
7. Rajput, M., Rajput A., Rajput, A. H.: Epidemiology. in *Handbook of Parkinson's disease*. R. Pahwa and K. E. Lyons, Eds., 4th ed., Informa Healthcare, (2007).
8. Sapir, S., Ramig, L., Spielman, J., Fox, C.: Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. in *Journal of Speech, Language and Hearing Research*, vol. 54, 114-125, (2010).
9. Szaszák, G., Beke, A.: Exploiting Prosody for Automatic Syntactic Phrase Boundary Detection in Speech. in *Journal of Language Modelling*, vol. 0, no. 1, 143-172, (2012).
10. Sztahó, D., Vicsi, K.: Estimating the Severity of Parkinson's Disease Using Voiced Ratio and Nonlinear Parameters. in *Proc. of 4th International Conference on Statistical Language and Speech Processing*, Pilsen, Czech Republic, 96-107, (2016).
11. Terapong B. at al: Assessment of Vocal Correlates of Clinical Depression in Female Subjects with Probabilistic Mixture Modeling of Speech Cepstrum. 11th International Conference on Control, Automation and Systems, (2011).
12. Thaweesak Y. at al: Characterizing Sub-Band Spectral Entropy Based Acoustics as Assessment of Vocal Correlate of Depression. International Conference on Control, Automation and Systems, (2010).
13. Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J., Ramig, L.O.: Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. in *IEEE Trans. on Biomedical Engineering*, vol.59, no.5, 1264-1271, (2012).
14. Szaszák, G., Tündik, M. A., Gerazov, B., Gjoreski, A.: Combining Atom Decomposition of the F0 Track and HMM-based Phonological Phrase Modelling for Robust Stress Detection in Speech. in *Proc. of 18th International Conference on Speech and Computer*, Budapest, Hungary, 165-173, (2016).
15. Vicsi, K., Sztahó, D.: Problems of the Automatic Emotion Recognitions in Spontaneous Speech; An Example for the Recognition in a Dispatcher Center. in *Esposito, A. et al.: Toward Autonomous, Adaptive, and Context-Aware Multimodal Interfaces. Theoretical and Practical Issues*. Heidelberg: Springer, 331-339, (2011).